

DNA sequencing leads to genomics progress in China

WU JiaYan[†], XIAO JingFa[†], ZHANG RuoSi & YU Jun^{*}

CAS Key Laboratory of Genome Science and Information, Beijing Institute of Genomics,
Chinese Academy of Sciences, Beijing 100029, China

Received January 6, 2011; accepted February 13, 2011

Citation: Wu J Y, Xiao J F, Zhang R S, *et al.* DNA sequencing leads to genomics progress in China. *Sci China Life Sci.* 2011, 54: 290–292, doi: 10.1007/s11427-011-4148-z

1 Science in the large-scale sequencing era

Ten years ago, the first draft sequence assembly of the human genome was completed [1], bringing biomedical research one-step closer toward the goal of revolutionizing diagnosis, prevention, and treatment of human diseases. Recently, journalists from the journal *Nature* surveyed more than 1000 life scientists regarding this laudable aim [2], obtaining substantially negative responses [3]. However, almost all of those surveyed had been influenced, in one way or another, by the availability of the human genome sequence, and they also agreed with the notion that the “sequence is the start.” The complexity of genome biology and almost every aspect of human biology is far greater than previously thought [4].

The Ministry of Science and Technology in China, and several local governments, have helped to establish several genomics institutions since 1998: the Chinese National Human Genome Center in Shanghai (commonly known as the Southern Center), the National Human Genome Center in Beijing (the Northern Center), the Beijing Institute of Genomics, Chinese Academy of Sciences (known as BIG-CAS; established initially as the Human Genome Center at the Institute of Genetics), and the Beijing Genomics Institute (also known as BGI or Huada, which had been part of BIG-CAS until it moved to Shenzhen) [5]. Chinese scientists contributed data to the International Human Genome

Project and the HapMap Project, as well as several other multinational genome projects. Genomic studies in China have paved the way for interdisciplinary research and collaboration in life sciences in China.

2 Next generation sequencing

Over the past 30 years, DNA sequencing technology has been one of the most important technological advances in biology. It started with capillary-based semi-automated implementations using the chemistry developed by Sanger [6–8]. This method achieved read-lengths up to 1000 bp with a cost of \$0.50 per kilobase, and a per-base accuracy of 99.999% [9]. Next generation sequencers (NGSs) have significantly reduced the cost of DNA sequencing, and enhanced the acquisition of a growing variety of molecular data, including, but not limited to, genetic variation, RNA expression, protein-DNA interaction, and chromosome structure change. The cost of DNA sequencing has been reduced by several orders of magnitude. More and more individual investigators can pursue projects on a scale previously accessible only to major genome centers. Roche (such as GS FLX), Illumina (such as HiSeq2000), and Life Technologies (such as SOLiD 5500x1) are the leading companies that manufacture major NGS platforms and share the current sequencing market.

Although the NGS sequencing strategies still produce shorter read lengths, the increasing availability of whole genome assemblies for *Homo sapiens* and major model or-

[†]Contributed equally to this work

^{*}Corresponding author (email: junyu@big.ac.cn)

ganisms has been tremendous. As shown in Figure 1, there were only 84 bacterial genome sequences available in 2002, compared to 1163 available now. The number of archaeal genome projects has grown five-fold in the past eight years (16 in 2002 vs. 93 in 2010). The increase is more striking for eukaryotic genome sequencing; from 2002 to 2010 the number of available genome sequences rose from 4 to 141 for fungi and from 8 to 151 for animals (Figure 2).

In the past 12 months, five large genome projects have been published: the domesticated apple [10], songbird [11], giant panda [12], cucumber [13], and maize [14]. The giant panda genome was finished using only Illumina's Genome Analyzer, which is the first whole genome assembly generated from just short sequencing reads, albeit with the obvious advantage of gene co-linearity of mammalian genomes.

NGS has brought many other changes in addition to a large amount of DNA sequences [15]. The rapidly falling sequencing cost suggests that we are not very far from the era of personalized genomics and medicine [16]. For example, based on the fast-accumulating human genetic data, researchers from diverse backgrounds are converging on the view that human evolution has been shaped by gene-culture interactions [17].

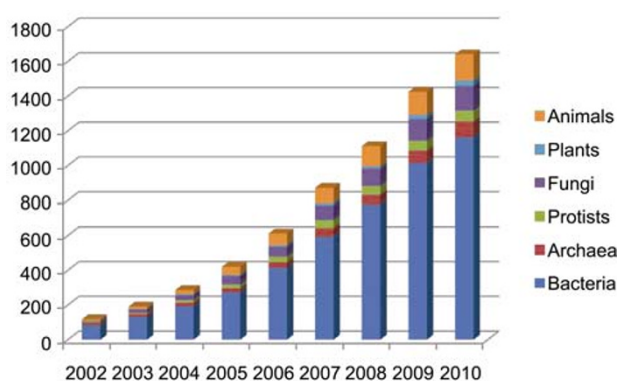


Figure 1 Whole genome projects completed between 2002 and 2010, including eukaryotes and prokaryotes.

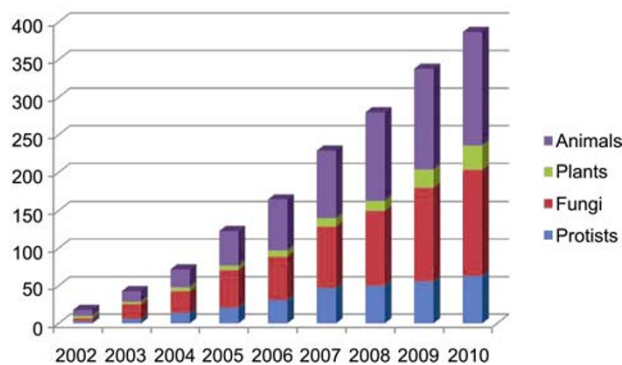


Figure 2 Eukaryotic whole genome projects completed between 2002 and 2010.

3 Genome projects in China

As Drs. Chen Zhu and Zhao Guoping mentioned in their recent publication [5], within the past ten years, the progress of genomics in China has been a great leap forward in terms of scale, quality, and inter-disciplinarily collaboration, both nationally and internationally. The intensive genomic research on human [18] and rice [19] genomes are unprecedented, allowing Chinese scientists to be the forerunners of these fields. For example, the Super-hybrid Rice Genome Project not only produced the first plant genome sequence using the whole-genome shotgun strategy, but also is still generating large amounts of sequence information in genetics and gene expression, which allows rice biologists to explore the molecular mechanisms of domestication and hybrid vigor [20,21]. By taking full advantage of the technology and resources of human genome research, Chinese life scientists have extended genomic sequencing and related research to many plants, microorganisms (pathogens for medicine and agriculture or important industry bacteria) [22–25], insects (silkworm), and parasites (*Schistosoma japonicum*), as shown as Table 1. In 2006, the assembled genomic sequence data of *S. japonicum* was registered and released from a public bioinformatics database (<http://biodb.sgst.cn>), operated by the Shanghai Bioinformatics Technology Development Center, for sharing with the international *Schistosoma mansoni* consortium [26].

4 Greater challenges ahead

The enormous advancements in sequencing technology have sent the cost of sequencing tumbling, and the daily

Table 1 Examples of genome projects completed by or heavily involving Chinese organizations

Species	Organization	Document
<i>Homo sapiens</i> (as part of HGP)	BIG-CAS	<i>Nature</i> 2001; 409: 860–921
<i>Oryza sativa</i> (indica cultivar-group) indica	BIG-CAS	<i>Science</i> 2002; 296(5565): 79–92
<i>Bombyx mori</i>	Southwest Agricultural University and BIG-CAS	<i>Science</i> 2004; 306(5703): 1937–40
<i>Gallus domesticus</i>	BIG-CAS	<i>Science</i> 2004; 432(7018): 695–716
<i>Homo sapiens</i>	BGI (Shenzhen)	<i>Nature</i> 2008; 456: 60–65
<i>Schistosoma japonicum</i>	Chinese Human Genome Center at Shanghai	<i>Nature</i> 2009; 460: 345–351
<i>Cucumis sativus</i> L.	BGI (Shenzhen)	<i>Nature Genetics</i> 2009; 41: 1275–1281
<i>Ailuropoda melano-leuca</i>	BGI (Shenzhen)	<i>Nature</i> 2010; 463: 311–317
<i>Harpegnathos saltator</i>	BGI (Shenzhen)	<i>Science</i> 2010; 329(5995): 1068–1071

output of sequencing operations worldwide have been soaring since 1993 [27]. Cheaper and faster sequencing, however, has brought new challenges. In bioinformatics and computational biology, the exponentially increasing amounts of sequence data have formed something of a backlog and biologists are not able to handle and mine the information as useful knowledge in their own disciplines. The lack of sharable software tools and user-friendly algorithms to analyze the genomic data, the shortage of qualified bioinformaticians, data storage/trafficking, and, to a lesser extent, raw computing power, form the most problematic issues for both funding agencies and researchers.

To revolutionize drug discovery, develop novel diagnostics and vaccines, and tailor disease treatment to an individual's genetic make-up; these are the most rational goals for genome sequencing and exploring human biology. In *Nature's* poll, more than one-third of respondents predicted that it will take 10–20 years for personalized medicine to be realized. More than 25% predicted that it will take even longer, but only 5% said that it will not happen in their lifetime. What do you think? More importantly, what could you do to make it happen sooner rather than later?

This work was supported by the Special Foundation Work Program (Grant No. 2009FY120100) and the National Key Technology Research and Development Program, Ministry of Science and Technology of China (Grant No. 2008BA164B02).

- 1 Macilwain C. World leaders heap praise on human genome landmark. *Nature*, 2000, 405: 983–984
- 2 Genome survey results, <http://go.nature.com/3Aywn>, *Nature*, Genome survey, 2010
- 3 Butler D. Science after the sequence. *Nature*, 2010, 465: 1000–1001
- 4 Check Hayden E. Human genome at ten: Life is complicated. *Nature*, 2010, 464: 664–667
- 5 Chen Z, Zhao G P. Human genomics in China—ten years endeavor: from planning to implementation. *Sci China Ser C-Life Sci*, 2009, 52: 2–6
- 6 F Sanger, G M Air, B G Barrell, et al. Nucleotide sequence of bacteriophage ϕ X174 DNA. *Nature*, 1977, 265: 687–695
- 7 Swerdlow H, Wu S L, Harke H, et al. Capillary gel electrophoresis

- for DNA sequencing. Laser-induced fluorescence detection with the sheath flow cuvette. *J Chromatogr*, 1990, 516: 61–67
- 8 Hunkapiller T, Kaiser R J, Koop B F, et al. Large-scale and automated DNA sequence determination. *Science*, 1991, 254: 59–67
- 9 Shendure J, Ji H L. Next-generation DNA sequencing. *Nat Biotechnol*, 2008, 26: 1135–1145
- 10 Velasco R, Zharkikh A, Affourtit J, et al. The genome of the domesticated apple (*Malus domestica* Borkh). *Nat Genet*, 2010, 42: 833–839
- 11 Warren W C, Clayton D F, Ellegren H, et al. The genome of a songbird. *Nature*, 2010, 464: 757–762
- 12 Li R, Fan W, Tian G, et al. The sequence and de novo assembly of the giant panda genome. *Nature*, 2010, 463: 311–317
- 13 Huang S, Li R, Zhang Z, et al. The genome of the cucumber, *Cucumis sativus* L. *Nat Genet*, 2009, 41: 1275–1281
- 14 Schnable P S, Ware D, Fulton R S, et al. The B73 maize genome: complexity, diversity, and dynamics. *Science*, 2009, 326: 1112–1115
- 15 Baker M. Next-generation sequencing: adjusting to data overload. *Nature Method*, 2010, 7: 495–499
- 16 Guttmacher A E, McGuire A L, Ponder B, et al. Personalized genomic information: preparing for the future of genetic medicine. *Nat Rev Genet*, 2010, 11: 161–165
- 17 Laland K N, Sean Myles J O-S. How culture shaped the human genome: bringing genetics and the human sciences together. *Nat Rev Genet*, 2010, 11: 137–148
- 18 Lander E S, Linton L M, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*, 2001, 409: 860–921
- 19 Yu J, Hu S, Wang J, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science*, 2002, 296: 79–92
- 20 Wei G, Tao Y, Liu G Z, et al. A transcriptomic analysis of superhybrid rice LYP9 and its parents. *Proc Natl Acad Sci USA*, 2009, 106: 7695–7701
- 21 Huang X, Wei X, Sang T, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet*, 2010, 42: 961–967
- 22 Wang L, Liu B, Zhou Z M. Research progress in genomics of environmental and industrial microorganisms. *Sci China Ser C-Life Sci*, 2009, 52: 64–73
- 23 Han Z G. Recent progress in genomic research of liver cancer. *Sci China Ser C-Life Sci*, 2009, 52: 24–30
- 24 Li H W, Li J Z, Zhao G P. Current research status of immunology in the genomic era. *Sci China Ser C-Life Sci*, 2009, 52: 43–49
- 25 Yang R F, Guo X K, Yang J. Genomic research for important pathogenic bacteria in China. *Sci China Ser C-Life Sci*, 2009, 52: 50–63
- 26 Zhou Y, Zheng H, Chen Y, et al. The *Schistosoma japonicum* genome reveals features of host–parasite interplay. *Nature*, 2009, 460: 345–351
- 27 Human genome at ten: The sequence explosion. *Nature*, 2010, 464: 670–671

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.